

THE INFORMATION CAPACITY OF NERVE CELLS USING A FREQUENCY CODE

RICHARD B. STEIN

From the University Laboratory of Physiology, Oxford, England

ABSTRACT Approximate equations are derived for the amount of information a nerve cell or group of nerve cells can transmit about a stimulus of a given duration using a frequency code (i.e., assuming the mean frequency of nerve impulses measures the intensity of a maintained stimulus). The equations take into account the variability of successive interspike intervals, and any serial correlations between successive intervals, but do not require detailed assumptions about the mechanism of impulse initiation. The errors involved in using these approximations are evaluated for neurons which discharge either completely regularly, completely at random (Poisson process) or show a particular type of intermediate variability (gamma distribution model). The errors become negligibly small as the stimulus duration or the number of functionally similar nerve cells increases. The conditions for applying these equations to experimental data are discussed. The application of these equations should help considerably in eliminating the enormous discrepancies between some earlier estimates for the information processing capabilities of single nerve cells and systems of nerve cells.

INTRODUCTION

Since Shannon (1948) developed a theory for transmission of information in electrical systems, biologists, engineers, and mathematicians have tried to apply this theory to the study of nerve cells. The theory was appealing since, at least in the periphery, one of the major functions of nerve cells is the transmission of information about the environment to the central nervous system. Thus, information theory offered the possibility of comparing quantitatively the function of different nerve cells and sensory systems. However, attempts to apply information theory have led to enormous discrepancies. Rapoport and Horvath (1960) estimated the information capacity of single cells at 4000 bits of information per second, but Jacobsen (1950, 1951), studying systems of cells, calculated capacities per cell of only 0.3 and 5 bits/sec. Finally, Quastler (1956) found that whole organisms containing 10^{10} or more nerve cells could transmit less than 100 bits/sec. In this paper I shall reexamine systematically the information capacity of individual nerve cells and groups of

cells and try to clarify the reasons for these discrepancies. In addition, a number of formulae will be derived which should greatly simplify calculations, from experimental data, of the information transmitted by nerve cells.¹

The discussion of information processing in the nervous system will follow closely Shannon's now famous development of the subject for other communications systems. However, there is a fundamental difference between the code used by many nerve cells and that used in most electrical communications systems. A common form of electrical coding is a binary pulse code. Time is divided into discrete intervals and in each interval one of two alternatives or one binary bit of information is sent. Since a neuron can fire up to 1000 all-or-none impulses/sec, it might transmit 1000 bits/sec with a binary code. Rapoport and Horvath (1960, see also Mackay and McCulloch, 1952) estimated that using the duration of each interval to encode information could increase transmission, possibly to 4000 bits/sec. These estimates are radically wrong in practice for at least two reasons. First, many neurons, particularly in the peripheral nervous system use neither a binary code nor an interval code, but a frequency code in which the stimulus intensity is signified by different numbers of nerve impulses over a period of time (Adrian, 1932). Successive interspike intervals may vary in duration when the stimulus is held constant, but the fluctuations are inherent in the impulse-generating process (Stein, 1967) and often, only the mean number of nerve impulses over a period of time, i.e. the mean impulse frequency, conveys information. Even in systems of fast-adapting tactile neurons where primary afferent fibers discharge only a single impulse (Armett et al., 1962), the second order neurons fire increasing numbers of impulses as the stimulus intensity is increased. Secondly, in the studies mentioned by Rapoport and Horvath, finely graded, carefully timed electrical stimuli were applied to axons. However, these studies are not applicable since in the normal synaptic activation process, pre-synaptic impulses produce larger, quantal effects at random times so that temporal fluctuations are orders of magnitude larger (Stein, 1965).

Barlow (1963) noted that for binaural localization of a sound source a certain amount of precise temporal information must be preserved, and Hall (1964) clarified the neuronal mechanisms of this phenomenon. However, this paper will concentrate on frequency coding because of its widespread occurrence peripherally and hence its importance in transmission of sensory information to the central nervous system. To distinguish eight different intensities, a frequency code in its simplest form would require stimuli capable of producing from zero to at least seven nerve impulses. Thus, frequency coding is inefficient (with a binary code, the occurrence or nonoccurrence of impulses in three intervals could specify eight intensities) and becomes progressively less efficient with increasing numbers of alternatives. However, the redundancy introduced by inefficient coding has definite advantages for a biological system, one of whose main functions is survival. With frequency coding, 6

¹ A brief account of this work was presented to the 2nd International Biophysics Congress which took place in Vienna on September 5-9, 1966.

impulses, if not referring to stimulus intensity 6, probably refers to intensity 5 or 7, but the binary code 110 = 6, if wrong, could have the quite different value 010 = 2.

These ideas can be made more precise. If one applies a stimulus to a nerve cell for a time t , a number j nerve impulses will be discharged. The number of impulses discharged on a particular trial will be considered as the output or response variable, and will be denoted y_j if j impulses are discharged. If the same stimulus is repeated, the number of impulses may vary, but from a large number of trials one can determine the mean number of nerve impulses x produced in the time t . If the "state" of the nerve cell is constant from trial to trial (e.g. there are no changes in the membrane potential, membrane resistance and the level of efferent control, if any, exerted by the nervous system on the cell's activity), the mean number will be determined by the stimulus alone and will often increase monotonically with the stimulus strength. The mean number of nerve impulses generated x can then be considered as the input or stimulus variable for the nerve cell, and the range of stimuli is defined by $x_{\min} \leq x \leq x_{\max}$ where x_{\min} and x_{\max} are the minimum and maximum mean numbers of impulses produced by natural stimuli.

In any natural or experimental situation, a variety of stimuli will be applied, and $p(x)$ will denote the probability that the stimulus is such that on average x impulses are generated. Any random feature in the process of impulse initiation will cause the number of impulses to vary from trial to trial, and there will be a set of conditional probabilities that j impulses are discharged, $j = 0, 1, 2, \dots \infty$, when the stimulus is adjusted so that the mean number is x .² These conditional probabilities will be denoted $p(y_j/x)$, and they specify the properties of a nerve cell as a communications channel. The over-all probability $p(y_j)$ of j impulses being discharged is obtained by integrating over all possible stimuli,

$$p(y_j) = \int_{x_{\min}}^{x_{\max}} p(y_j/x) p(x) dx \quad (1)$$

These relationships are illustrated in Fig. 1.

Central to the formulation of information theory is a measure for the uncertainty of a probability distribution. From the standard formulae (Ash, 1965) one can define an uncertainty $H(Y)$ for the response distribution

$$H(Y) \equiv - \sum_{j=0}^{\infty} p(y_j) \log p(y_j) \quad (2)$$

and a conditional uncertainty $H(Y/x)$ in the set of responses Y given the stimulus x

$$H(Y/x) \equiv - \sum_{j=0}^{\infty} p(y_j/x) \log p(y_j/x) \quad (3)$$

² Although much of the paper will be concerned with information transmission by single nerve cells, these definitions and many of the results are easily extended to functionally similar groups of cells (see section 2 *h*).

An important quantity is the average conditional uncertainty $H(Y/X)$ over the entire stimulus set X

$$H(Y/X) \equiv \int_{x_{\min}}^{x_{\max}} p(x) H(Y/x) dx \quad (4)$$

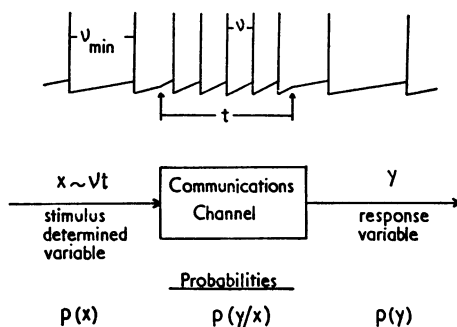


FIGURE 1 The top line shows schematically a neuronal discharge having a steady frequency v_{\min} . The stimulus of duration t increases the frequency and produces x impulses on average. This is often considered as the neuron's response, but it is treated here as the input to a communications channel, since the average frequency or mean number of impulses may be the only quantity determined wholly by the stimulus. The response variable y is assumed to be the number of nerve impulses actually discharged on a single trial. This will vary from trial to trial because of random fluctuations in the part of the neuronal cycle at which the stimulus begins, in the duration of successive interspike intervals, and in the "state" of the neuron. This variability limits the amount of information about the stimulus that the nervous system (or the experimenter) can obtain from knowing the number of nerve impulses discharged on a single trial. The amount of information transmitted depends on three probability density functions: $p(x)$, the probability that the stimulus on a trial is x ; $p(y)$, the probability that the response is y ; $p(y/x)$, the probability that the response is y given that the stimulus is x . The last probability density function determines the properties of a neuron as a communications channel. From it alone one can determine both the capacity of a neuron to transmit information and the stimulus distribution necessary to attain the information capacity.

The information transmitted is the difference

$$I \equiv H(Y) - H(Y/X) \quad (5)$$

and the (neuronal) information capacity is

$$C \equiv \max_{p(x)} I. \quad (6)$$

I shall not attempt to derive these formulae here, but only mention some points which are essential in understanding the application of these formulae to nerve cells.

(a) X and Y denote the set of inputs and outputs (stimuli and responses) respec-

tively. For particular stimuli or responses, small letters (x and y) are used, and these will be given subscripts when they can take only certain, discrete values. Since the stimulus is continuously variable and the response is discrete, a mixture of integrals and sums is necessary in equations (2)–(4). In most other applications the input and output variables are both discrete or both continuous and examples of these types of channels will also be considered.

(b) The form of $H(Y/X)$ indicates that it is determined by two sets of probabilities, $p(x)$ which depends only on the stimuli, and $p(y_j/x)$ which depends only on the properties of the neuronal information channel. $H(Y)$ is determined by these same sets of probabilities as indicated by equations (1) and (2). These two sets of probabilities determine the information transmission, and both can be determined experimentally without reference to the detailed way in which the stimuli are transformed and nerve impulses initiated. Thus, the formulae offer a framework for comparing experimentally different nerve cells, provided one can determine the appropriate range of biological stimuli and make assumptions about the coding used.

(c) The term response uncertainty for $H(Y)$ should be clear since $H(Y)$ depends only on the probability of the various responses. Similarly in equation (4), the average conditional uncertainty $H(Y/X)$ measures the average uncertainty that remains, even if the stimuli are specified. The information transmitted, equation (5), measures the reduction in uncertainty that results from consideration of the correlation between stimuli and responses. To obtain the information capacity C , one must consider all possible stimulus distributions so as to find the one which gives the maximum information transmission.

(d) If logarithms to the base 2 are used, the values of uncertainty, information, and information capacity (see also Ash, 1965) will be measured in bits per stimulus. The values of capacity quoted earlier had been converted to bits/second by assuming a particular stimulus repetition rate. If the responses to successive stimuli are independent, this conversion presents no new difficulties. If not, the particular type of dependence must be considered. The results of this paper will be limited to independent stimuli though the problem of interaction between successive stimuli will be considered briefly in the Discussion.

The number of categories M that can be distinguished from the set of responses represents the antilog of an informational measure. If the information of equation (5) is measured in bits, then

$$M = 2^I \quad (7)$$

Natural logarithms are more convenient in deriving analytical results. Where confusion between the two systems is possible, the subscripts 2 and e have been used. To convert numerical values for uncertainty in natural units to bits, one multiplies by $\log_2 e$.

(e) In an electrical communications system, by construction of an efficient en-

coder, the information capacity can always be approached in principle with sufficiently long messages. However, the biological organism may not be able to alter sufficiently the way in which the natural stimuli are encoded to approach the information capacity. Nonetheless, the concept of information capacity is a useful one, and it would be interesting to investigate to what extent natural selection has matched the properties of peripheral nerve cells to the distribution of stimuli that they normally receive.

The amount of calculation and experimentation involved, particularly if one wishes to determine the information capacity directly, has severely limited the number of attempts to apply information theory experimentally to nerve cells, though there are some notable examples in the literature (FitzHugh, 1957; Werner and Mountcastle, 1965). The present results are divided into two sections. First, some specific examples of practical interest are considered to illustrate the use and simplification of the equations. In the second section, general results are derived for the limit where a maintained stimulus produces large numbers of responses. Details of proofs which are not essential to a general understanding of the paper have been included in small type.

RESULTS

1. *Some Examples*

A general analytic expression for the information capacity of a neuron using a frequency code is not possible for two reasons. First, there is at present no adequate neuronal model from which the conditional probabilities, $p(y_i/x)$, can be derived in closed form. Secondly, even for quite simple models, the calculation of information capacity rapidly becomes involved enough to require considerable computer time. However, reasonably simple results can be obtained with certain assumptions. Although the experimental neurophysiologist may find the initial examples extreme, I shall move continually toward more realistic but analytically more difficult examples. The application of these results to experimental data will be discussed.

(a) *Single Nerve Action Potentials (the Binary Channel)*. The simplest information channel is a binary channel in which there are two possible responses. I shall call the response y_1 if no impulse is discharged and y_2 if there are one or more impulses.³ Assume as well two possible stimuli, x_1 and x_2 , which have probabilities a and b respectively of generating an impulse where, for definiteness $b > a$. These properties can be specified by a transition or channel matrix.

$$P = \begin{bmatrix} 1 - a & a \\ 1 - b & b \end{bmatrix}$$

³ The notation for sections (1 a) and (1 b) differs slightly from that outlined in the Introduction. If there are no impulses, the response is denoted y_1 not y_0 to conform to standard matrix notation in which indices start at one.

whose elements are $p_{ij} = p(y_j/x_i)$; i.e., the element in the i th row and j th column gives the probability that y_j is the response to stimulus x_i . To maximize information transmission a and b must differ by maximum amounts ($x_1 = x_{\min}$ and $x_2 = x_{\max}$) and the relative probabilities of occurrence of stimuli x_1 and x_2 must be optimal. If $a = 0$ and $b = 1$, one bit can be transmitted. For other fixed values of a and b , the channel capacity in bits, C_2 , is

$$C_2 = (-bH_a + aH_b)/(b - a) + \log_2 [1 + 2^{(H_a - H_b)/(b-a)}] \quad (1.1)$$

where

$$\begin{aligned} H_a &\equiv H(Y/x_1) \equiv - \sum_{j=1}^2 p(y_j/x_1) \log_2 p(y_j/x_1) \\ &= - (1 - a) \log_2 (1 - a) - a \log_2 a \end{aligned}$$

and

$$H_b \equiv H(Y/x_2) = - (1 - b) \log_2 (1 - b) - b \log_2 b$$

To derive this formula, one must invert the channel matrix above which gives the matrix

$$Q = (b - a)^{-1} \begin{bmatrix} b & -a \\ b - 1 & 1 - a \end{bmatrix}$$

Using the elements q_{ij} of this inverted matrix, one can compute the information capacity from a general equation for a channel with m inputs and outputs (Ash, 1965)

$$C_e = \log_e \sum_{j=1}^m \exp \left[- \sum_{i=1}^m q_{ji} H(Y/x_i) \right] \quad (1.2)$$

where $H(Y/x_i) = - \sum_{j=1}^m (p(y_j/x_i) \log_e p(y_j/x_i))$. The exponential is used with natural logarithms and capacity in natural units. With logarithms to the base 2, the expression in square brackets becomes the exponent of 2. Substituting for the binary channel, one obtains (see Ash, 1965)

$$C_e = \log_e \left(\exp \left[\frac{-bH_a + aH_b}{b - a} \right] + \exp \left[\frac{(1 - b)H_a + (a - 1)H_b}{b - a} \right] \right)$$

which reduces to equation (1.1) above.

If there is no spontaneous activity ($a = 0$), $H_a = 0$, so the first term on the right hand side of equation (1.1) is zero and the channel capacity simplifies to

$$C = \log [1 + b(1 - b)^{(1-b)/b}] \quad (1.3)$$

The threshold stimulus is often defined as that stimulus that elicits a response 50%

of the time ($b = \frac{1}{2}$). Then, from equation (1.3) only $\log_2 (1\frac{1}{4}) = 0.32$ bits of information could be transmitted per stimulus. (FitzHugh, 1957, has compared in more detail threshold and informational measures.)

Another point of interest is the behavior of the information capacity for stimuli short enough that the probability of a response is much less than one. If a cell is discharging impulses spontaneously at a frequency ν_{\min} , as illustrated in Fig. 1, there will be a probability $a = \nu_{\min} t$ of an impulse in a short time t . This probability may be increased to $b = \nu_{\max} t$ when the maximum stimulus is applied, but if a and b are much less than one, then the capacity reduces to

$$C \sim \nu_{\max} t \left[\frac{\log e}{e} \left(\frac{\nu_{\max}}{\nu_{\min}} \right)^k - k \log \left(\frac{\nu_{\max}}{\nu_{\min}} \right) \right] \quad (1.4)$$

where $k = \nu_{\min}/(\nu_{\max} - \nu_{\min})$.

To derive equation (1.4) it is best to use natural logarithms. Then, $H_a = -(1-a) \log_e (1-a) - a \log_e a \sim a(1 - \log_e a)$, if $a \ll 1$. Similarly, $H_b \sim b(1 - \log_e b)$, if $b \ll 1$, so that the first term on the right hand side of equation (1.1) becomes

$$\frac{-bH_a + aH_b}{b-a} \sim \frac{ab(\log_e a - \log_e b)}{b-a} = \frac{-\nu_{\min} \nu_{\max} t \log_e (\nu_{\max}/\nu_{\min})}{\nu_{\max} - \nu_{\min}} \quad (1.5)$$

Also,

$$\frac{H_a - H_b}{b-a} \sim -1 + \frac{b}{b-a} \log_e b - \frac{a}{b-a} \log_e a = \log_e \left[\left(\frac{\nu_{\max}}{\nu_{\min}} \right)^k \frac{\nu_{\max} t}{e} \right]$$

Thus, for the second term on the right hand side of equation (1.1),

$$\log_e \left[1 + \exp \left(\frac{H_a - H_b}{b-a} \right) \right] \sim \left(\frac{\nu_{\max}}{\nu_{\min}} \right)^k \frac{\nu_{\max} t}{e} \quad (1.6)$$

Combining equation (1.5) and (1.6) produces equation (1.4). In equation (1.4) the factor $\log e$ has been inserted so any logarithmic base can be used, e.g., $\log_2 e = 1.443$, $\log_{10} e = 0.4343$.

As a typical example, let $\nu_{\max} = 100$ impulses/sec and $\nu_{\min} = 10$ impulses/sec. These values will be used throughout the paper so the results that follow from different assumptions will be comparable. Then, $k = \frac{1}{9}$ and equation (1.4) becomes $C \sim 32t$. Thus, the amount of information one can send along a neuronal channel initially increases linearly with time, though the regression coefficient 32 is less than one might expect with a total range of $100 - 10 = 90$ impulses/sec. As time increases, the information capacity depends on neuronal variability, and I shall now consider the effects of making different assumptions about neuronal variability.

(b) *A Completely Regular Neuron.* Consider a model neuron such as that illustrated in Fig. 1. Depolarization proceeds linearly until threshold. Then, after a

nerve spike which is assumed short compared to the ramp depolarization duration, the depolarization is erased and the cycle is repeated. A stimulus is assumed merely to alter the rate of depolarization. Calvin and Stevens (1966) used a model similar to this to explain some properties of cat motoneurons. If stimuli of duration t are applied at random to such a regularly firing cell, the stimulus will occur at different points of the neuronal cycle. The probability that the stimulus generates a nerve impulse will then increase linearly with stimulus duration. With a maximal stimulus, the probability will reach one at $t = \mu_0 = 1/\nu_{\max}$ where ν_{\max} is the minimum mean interspike interval. Similarly, there will be a maximum mean interspike interval $\mu_1 = 1/\nu_{\min}$, if $\nu_{\min} > 0$. For the example cited above, the information transmission rate increases with stimulus duration until at $\mu_0 = 10$ msec, the information capacity calculated from equation (1.1) is 0.76 bits (compared to $32 \times 0.01 = 0.32$ bits which the results above for short times would indicate). If the duration is in the range $\mu_0 < t < 2\mu_0$, a maximal stimulus will sometimes produce two impulses. One can also choose a third intermediate stimulus that will always produce one and only one impulse.

The channel matrix is thus a 3×3 matrix as follows:

$$\begin{bmatrix} 1 - a & a & 0 \\ 0 & 1 & 0 \\ 0 & 1 - b & b \end{bmatrix}$$

where the three stimuli x_1 , x_2 , and x_3 are the minimum, intermediate, and maximum stimuli respectively, while the three responses y_1 , y_2 , and y_3 indicate the occurrence of 0, 1, and 2 nerve impulses. The matrix element p_{ij} is the conditional probability $p(y_j/x_i)$ as before, and it is assumed that $a = \nu_{\min} t$ and $b = \nu_{\max} t - 1$.

In the region $2\mu_0 < t < 3\mu_0$ sometimes 3 impulses occur in response to a maximal stimulus. One could then select two intermediate stimuli that always produced 1 and 2 impulses respectively. The channel matrix would be

$$\begin{bmatrix} 1 - a & a & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 - b & b \end{bmatrix}$$

where all quantities are as previously defined except that x_4 is now the maximal stimulus and $b = \nu_{\max} t - 2$. It is an important property of this matrix that there are subsets of stimuli that are never confused. Stimuli x_1 and x_2 never produce more than one impulse (response y_2), while x_3 and x_4 never produce less than two impulses. Now Shannon (1948) indicated that if the information capacity of the n th distinct subset is C_n when symbols (stimuli) in that subset alone are used, the total capacity is

$$C = \log_2 \sum_n 2^{C_n} \quad (1.7)$$

The matrices for the subset in the present example are 2×2 matrices whose capacity is given by equation (1.1). Higher order matrices, defined as above for the regularly firing neuron, can be similarly subdivided. Fig. 2 shows calculated values for information capacity as a function of time for the first 100 msec assuming $\nu_{\max} = 100$ and $\nu_{\min} = 10$ impulses/sec. The capacity has periodicities of $\mu_0 = 10$ msec and $\mu_1 = 100$ msec, although only the first is seen in Fig. 2. The dashed lines in Fig. 2 are upper and lower limits for the information capacity, given by

$$\log [d + (\nu_{\max} - \nu_{\min})t] \leq C \leq \log [1 + (\nu_{\max} - \nu_{\min})t] \quad (1.8)$$

where d is a constant having a value of 0.42 in Fig. 2. The upper limit is an extension of a formula first proposed as an exact expression by Barlow (1963) for a neuron without spontaneous activity.

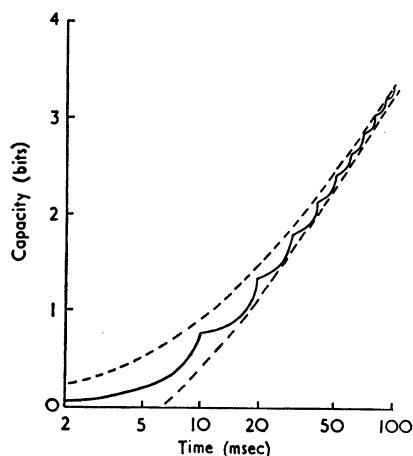


FIGURE 2 Information capacity of a completely regular neuron (solid line) as a function of the duration of a maintained stimulus. The dashed lines are upper and lower limits which converge rapidly as time (on a logarithmic scale) increases. The values were calculated for the example described in the text. The range of neuronal impulse frequencies was from 10 to 100 impulses/sec.

The upper limit is obtained as follows: if $\nu_{\max} t$ and $\nu_{\min} t$ are both integers, then there will be exactly $1 + (\nu_{\max} - \nu_{\min})t$ stimuli that can be applied without confusion and hence $\log_2 [1 + (\nu_{\max} - \nu_{\min})t]$ bits of information can be transmitted per stimulus. If either $\nu_{\max} t$ (or $\nu_{\min} t$) are not integers, there will be some confusion between the two highest (or lowest) stimuli and one can show that the capacity will be decreased below that given by the right-hand side of equation (1.8).

The value of d in the lower limit is more difficult to obtain directly. If there is no spontaneous activity, one can write $\nu_{\max} t = j + b$ where j is an integer and $0 \leq b < 1$. Then from equations (1.3) and (1.7) it follows that

$$C = \log [j + 1 + b(1 - b)^{(1-b)/b}]$$

One can show graphically or analytically that

$$1 + b(1 - b)^{(1-b)/b} \geq b + 0.71$$

for $0 \leq b \leq 1$, and thus $d = 0.71$. If spontaneous activity is present and $(\nu_{\max} - \nu_{\min})t > 2$, a similar argument can be used for the lowest possible numbers of nerve impulses and one finds that d is then $0.71 + (0.71 - 1) = 0.42$. It is this value of d that was used in Fig. 2, though it should be noted that the argument depends on being able to subdivide the channel matrix and consider upper and lower numbers of impulses separately. This is not always possible for $(\nu_{\max} - \nu_{\min})t < 2$, so a slightly smaller value of d may sometimes be necessary.

It follows easily from (1.8) that the limits converge as $1/t$ since for large t

$$\begin{aligned} \log [1 + (\nu_{\max} - \nu_{\min})t] - \log [d + (\nu_{\max} - \nu_{\min})t] \\ = \log \left[1 + \frac{1}{(\nu_{\max} - \nu_{\min})t} \right] - \log \left[1 + \frac{d}{(\nu_{\max} - \nu_{\min})t} \right] \\ \sim \frac{(1 - d) \log e}{(\nu_{\max} - \nu_{\min})t} \end{aligned} \quad (1.9)$$

The neuron was assumed to fire completely regularly and the only probabilistic aspect is the part of the cycle at which the stimulus begins. The deviation from the upper limit is due to the fact that the minimum and maximum stimuli may not be integers (all the intermediate stimuli were selected at integer values of x , and produce one and only one number of nerve impulses). Equation (1.9) simply makes precise the intuitive idea that the wider the stimulus range, the less important are the "end" values.

There is a further, perhaps less intuitive, result that follows from this analysis. I have not yet considered the distribution of stimulus probabilities that is necessary to produce the channel capacity. If the channel matrix is square, an exact expression is known (Ash, 1965, p. 56) and several optimal distributions are shown in Fig. 3 for different values of t . For a completely regular neuron an optimal stimulus distribution consists of a series of regularly spaced, equally probable stimuli, except at the very ends of the stimulus range. It is extremely unlikely that these particular stimulus distributions would occur naturally, and furthermore, these particular stimuli would not be optimal for other neurons with slightly different firing rates. It is thus of interest to consider not a discrete distribution of stimuli, but a continuous rectangular distribution, given by

$$\begin{aligned} p(x) &= (x_{\max} - x_{\min})^{-1}, & x_{\min} \leq x \leq x_{\max} \\ &= 0, & \text{otherwise.} \end{aligned} \quad (1.10)$$

If the stimulus range is large ($x_{\max} - x_{\min} \gg 1$), approximately $\frac{1}{2} \log_2 e = 0.72$ bits less information can be transmitted using this distribution than with the optimal discrete distribution.

This can be shown as follows: if the stimulus distribution, $p(x)$, is rectangular, the distribution of numbers of nerve impulses, $p(y)$, will be discrete, but flat except at the very ends of

the response range. If the frequency of nerve impulses remains constant, the range of numbers of impulses increases linearly with stimulus duration. Thus, as shown previously, "end effects" die out as $1/t$ and will be neglected. The response uncertainty is given by

$$H(Y) = - \sum_j p(y_j) \log p(y_j) \sim - \sum_j \frac{1}{M} \log \left(\frac{1}{M} \right) = \log M \quad (1.11)$$

where $M = (\nu_{\max} - \nu_{\min}) t$.

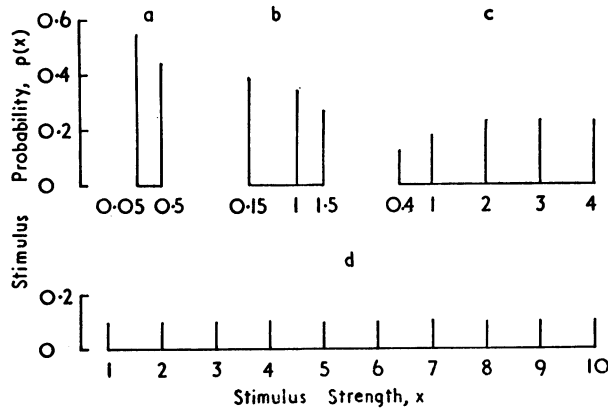


FIGURE 3 Stimulus distributions required to attain the information capacity of the completely regular neuron described in the text. The distributions consist of regularly spaced equiprobable stimuli (except toward the end of the stimulus range when the minimum or maximum stimulus does not produce an integral number of impulses). The stimulus durations are (a) 5 msec, (b) 15 msec, (c) 40 msec, and (d) 100 msec. With continuous stimulus distributions, less information could be transmitted.

From equation (4) the average conditional uncertainty is $H(Y/X) = \int p(x)H(Y/x) dx$ where $H(Y/x) = - \sum p(y_j/x) \log p(y_j/x)$. If one considers the unit segment in the middle of the stimulus continuum where $i < x < i + 1$, then

$$\begin{aligned} p(y_j/x) &= 0 & j < i \\ &= -x + i + 1, & j = i \\ &= x - i, & j = i + 1 \\ &= 0, & j > i + 1. \end{aligned}$$

Integrating from $x = i$ to $x = i + 1$,

$$\begin{aligned} \int_i^{i+1} p(x)H(Y/x) dx &= - \frac{1}{M} \left[\int_i^{i+1} (-x + i + 1) \log (-x + i + 1) dx \right. \\ &\quad \left. + \int_i^{i+1} (x - i) \log (x - i) dx \right] \quad (1.12) \end{aligned}$$

Substituting $z = -x + i + 1$ and using natural logarithms, the first integral on the right

hand side of (1.12) becomes,

$$\int_0^1 z \log_e z \, dz = -\frac{1}{4}$$

Similarly, substituting $z = x - i$, the second integral on the right hand side of (1.12) also integrates to $-\frac{1}{4}$ and the process can be repeated for each of the M unit segments of the stimulus continuum. Thus, except for end effects, $H(Y/X) \sim \frac{1}{2}$. The information transmitted in natural units is then $I_e = H(Y) - H(Y/X) \sim \log_e [(\nu_{\max} - \nu_{\min}) t] - \frac{1}{2} \sim C_e - \frac{1}{2}$. To convert to binary bits, one must multiply by $\log_2 e$ as before.

Although a decrease of a constant amount of information/cell of less than one bit may not seem important, it does mean a constant decrease of nearly 40% in the number of discriminable categories, since from equation (7)

$$I \sim \log M - \frac{1}{2} \log e = \log (M/\sqrt{e}) = \log (0.61M) \quad (1.13)$$

The effect of using continuous stimulus distributions was not taken into account by previous analyses (Grüsser, 1962; Barlow, 1963), nor has previous work attempted to analyze the effects of neuronal variability on information transmission via a frequency code. Variability is of course always present and later a criterion (2.14) will be given for deciding if the discharge is sufficiently regular for the formulae of this section to apply. Now, I shall turn to the opposite extreme, the randomly discharging neuron.

(c) *A Completely Random Nerve Cell (Poisson Process)*. Assume that the probability of j impulses occurring in response to a stimulus which produces x on average is given by a Poisson probability density function; i.e.,

$$p(y_j/x) = x^j e^{-x} / j! \quad (1.14)$$

This might occur in practice with one or more cells discharging randomly (Biscoe and Taylor, 1963), but superposition of a large number of processes will tend to produce a Poisson process over short periods of time, even if the individual processes are not of this type (Cox and Miller, 1965, p. 363). I shall generally refer to a single randomly firing cell, but the results will apply to the other examples as well.

The probability of no impulses is e^{-x} , so the probability of at least one impulse is $1 - e^{-x}$. Thus, one can consider a binary channel such as discussed previously, with $a = 1 - e^{-x} \sim x$, if x is small. For a neuron firing with frequency ν , $x = \nu t$ may grow linearly, but a will only slowly approach one. Thus, the channel capacity of a randomly discharging cell using a frequency code will be less than that of a regularly discharging one, and the difference increases with increasing stimulus duration. The channel capacity of a randomly discharging cell is illustrated in Fig. 4 using the same values as before, $\nu_{\max} = 100$ and $\nu_{\min} = 10$ impulses/sec. As the stimulus

duration increases, multiple responses occur more often, and at some stimulus duration, inclusion of a third, intermediate stimulus increases information transmission. Fig. 4 shows results calculated for binary, tertiary and quaternary channels ($m = 2, 3, 4$), together with the over-all information capacity (top solid line).

The method of calculation was as follows: for a given stimulus set, all input probabilities were initially made equal. The value of I was calculated for this probability distribution from equations (1)–(5) and (1.14). Then, each input probability was changed systematically by preset steps in the direction which increased the information until a maximum was reached. The step size was then reduced, a new maximum was found and the process was repeated until the maximum information was determined to within 0.001 bit, and the input probabilities required to produce this maximum were determined generally to an accuracy of 0.002. Since

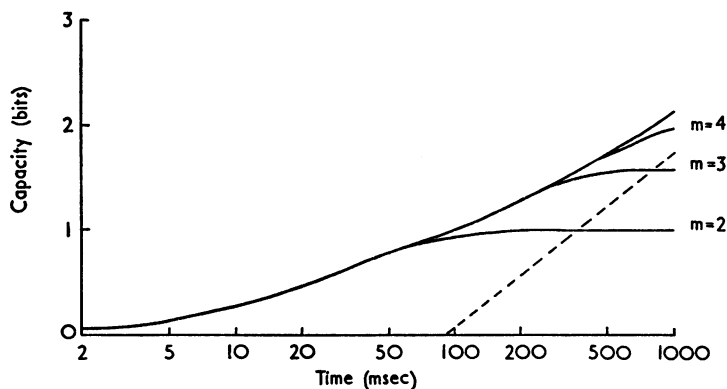


FIGURE 4 Information capacity of a randomly discharging neuron as a function of the duration of a maintained stimulus. The capacity using $m = 2, 3$, or 4 stimuli is indicated. The solid line extending above these is the over-all capacity using the optimal number of stimuli. The same range of mean impulse frequencies was used as for the regularly discharging neuron of Fig. 2. The time scale is logarithmic.

information is a convex function of the input probabilities (Ash, 1965), any maximum found in this way must be the capacity for that stimulus set (there are no local maxima).

To find the over-all neuronal information capacity a second level of iteration was required. Each of the stimulus values (x_i) was altered in turn, and the new channel capacity was calculated and compared with previous values. The process was continued until an over-all maximum was found. In varying the x_i , care must be taken to avoid local maxima which do occur. Generally, the maximum capacity for a given number of stimuli, each of which has a nonzero probability, occurred when the stimuli were approximately an equal number of standard deviation units apart (with a stimulus x , the standard deviation in number of impulses produced by a Poisson process is \sqrt{x}). Therefore, initial stimulus sets were selected which satisfied this criterion. This selection also reduced the requirements in computer time considerably. Nonetheless, determination of the over-all neuronal information capacity for some of the longer values of t shown in Fig. 4 took several minutes of computation using an ALGOL program on the English Electric KDF9 computer of the Oxford University Computing Laboratory.

The dashed line on the right is the simple formula

$$C = \frac{1}{2} \log (t/T) \quad (1.15)$$

where $T = 91.2$ msec. This is the asymptotic information capacity in the limit of large t . For $t = 1$ sec, the deviation from the asymptotic formula is 0.4 bits. (For N similar cells, the numbers of impulses will be multiplied by N and this degree of approximation will be reached at $1/N$ seconds.) Fig. 5 shows some of the stimulus distributions that are required to obtain the maximum channel capacity. As for the regularly firing neuron, the distributions consist (except near the ends of the stimulus range) of a set of nearly equiprobable stimuli, though the stimuli are no longer

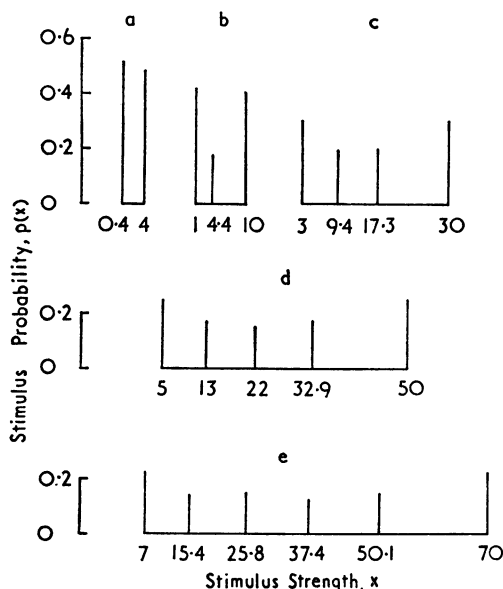


FIGURE 5 Stimulus distributions required to attain the information capacity illustrated in Fig. 4 for the randomly discharging neuron. The distributions consist, except at the ends of the stimulus range, of approximately equiprobable stimuli with the strengths indicated. The stimulus durations are (a) 40 msec, (b) 100 msec, (c) 300 msec, (d) 500 msec, and (e) 700 msec. The optimal stimulus spacing increases with stimulus strength, because of the increase in the standard deviation in number of impulses produced by a given stimulus (stimulus strengths are measured in terms of the number of impulses they produce). The optimal strengths are exact in (a), and were calculated to the nearest $\frac{1}{8}$ impulse in (b) to (d), and to the nearest $\frac{1}{2}$ impulse in (e). The difference between information transmission using these discrete distributions and related continuous distributions goes to zero as the stimulus duration is increased.

evenly spaced. The optimal distribution requires that the spacing be smaller at the lower stimulus values where the standard deviation in numbers of impulses is smaller. These discrete stimulus distributions are extremely unlikely under natural conditions, but I will show in section (2 b) that the difference between the information transmitted using a discrete distribution and a related continuous distribution approaches zero as t increases. This is in contrast to the result derived for the completely regular neuron, where the difference approached a constant, 0.72 bits.

(d) *Intermediate Variability (Gamma Distribution).* Most nerve cells do not fire either completely regularly or completely at random. To produce intermediate levels of variability, one can assume that a nerve cell produces an action potential only for every r th input (a typical input may be a presynaptic action po-

tential). If the inputs occur at random, the time to the r th input (interspike interval) obeys a gamma distribution. Even if one allows the effect of each input to decay exponentially, the interspike interval distributions can still be approximately fitted by a gamma distribution (Stein, 1965). The channel capacity was calculated for $r = 10$ (Fig. 6 a) and $r = 100$ (Fig. 6 b) with $\nu_{\max} = 100$ impulses/sec and $\nu_{\min} = 10$ impulses/sec as before. One can easily show that the coefficient of variation (the ratio of the standard deviation to the mean spike interval; see Stein, 1967) for a gamma distribution is $1/\sqrt{r}$ which equals 0.316 and 0.1 if $r = 10$ and 100 respectively.

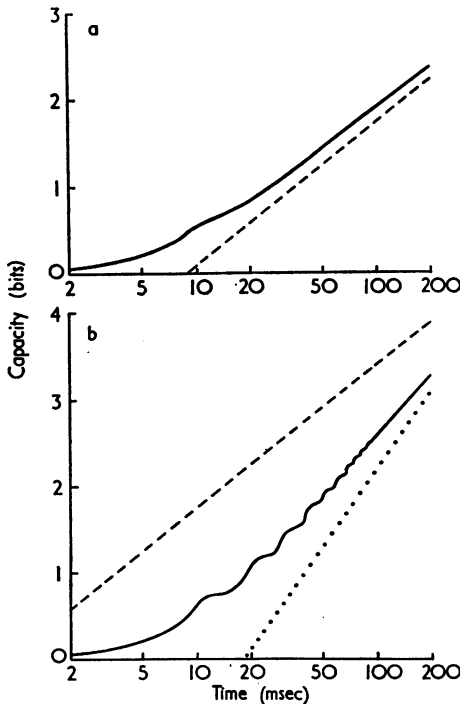


FIGURE 6 Information capacity of a neuron of intermediate variability. The values were calculated using a "gamma distribution" model in which r randomly occurring presynaptic impulses are required to produce a nerve impulse. In the upper graph $r = 10$ and in the lower one $r = 100$. As r increases the capacity at short durations approaches more closely that of a completely regular neuron. At long durations the slope of the capacity versus $\log t$ plot is a constant. Further explanation in text.

tively. Nerve cells which fire even more regularly are known, but the trends are clear from these two examples.

The method of computation was similar to that for the Poisson process described previously except for the values in the channel matrix. If $p(y_j/x)$ is the probability that j impulses are discharged in time t when the mean number is x , then

$$p(y_j/x) = \sum_{k=0}^{r-1} \binom{r-k}{r} \frac{(rx)^{(jr-k)} e^{-rx}}{(jr-k)!} + \sum_{k=1}^{r-1} \binom{k}{r} \frac{(rx)^{(jr+k)} e^{-rx}}{(jr+k)!}$$

The terms in the first sum are the products of two factors: the probability that at the start of the stimulus at least k events had already occurred since the last nerve impulse, and the probability that $jr - k$ events occur in time t from a Poisson process which produces rx on average.

Similarly, the terms in the second sum are the products of the probabilities that less than k events had already occurred since the last action potential before the start of the stimulus, and that $jr + k$ events occur during the time t when the mean number is rx (by summing over all j , one can verify that the mean number of nerve impulses will in fact equal x).

The channel capacity approaches the values for the regularly firing neuron (Fig. 2) more closely, the smaller the value of t or the higher the value of r . However, at long times the slope of the capacity versus $\log t$ plot approaches the same asymptotic value (indicated by dashed lines in Fig. 6) as the Poisson process, though shifted to the left on the time axis. A simple equation, which will be derived later, describes the asymptotic results

$$C = \frac{1}{2} \log (rt/T) \quad (1.16)$$

where $T = 91.2$ msec in this example. With $r = 10$, the information capacity approaches the asymptotic values closely (to within 0.2 bits) by 50 msec, but with $r = 100$, a difference of over 0.6 bits is still present at 200 msec. With higher values of r the approach to the asymptotic values will be even slower, but by retaining an extra term, a much more accurate asymptotic formula (2.15) can be derived. This is indicated by the dotted line in Fig. 6 *b* which is accurate to within 0.2 bits by 200 msec.

2. Some General Results

In the previous section, the capacity of nerve cells to transmit information using a frequency code was calculated for several important examples. A considerable amount of digital computer time was required to obtain exact values, though some simple formulae were mentioned which become increasingly accurate as the stimulus duration t is increased. I shall now derive these formulae as special examples of much more general asymptotic formulae. The basic proof depends on the assumption that (i) the discharge pattern is stationary and (ii) the variability in numbers of impulses is large enough so that the equation for the conditional uncertainty of a normal distribution can be used, but small enough relative to the total range of numbers of impulses that the response distribution is little different from the stimulus distribution. (Remember that the stimulus is defined in terms of the mean number of impulses it generates.) Sections (2 *a*) and (2 *b*) consider the two parts of the second assumption. Both become exact as $t \rightarrow \infty$ and this permits the derivation of the basic asymptotic results in section (2 *c*). However, the approach to the asymptotic formulae may be slow if the neuronal discharge is very regular; a more accurate approximation for this case is calculated in section (2 *d*). The final sections generalize the results for nonoptimal stimulus distributions, nonstationary discharge patterns, and populations of nerve cells.

(*a*) *Conditional Uncertainty.* Cox and Miller (1965, p. 343) prove an important result based on the central limit theorem of mathematics. It can be stated as follows: if the intervals between successive nerve impulses are independent,

random variables (renewal process) whose probability distribution has mean μ and variance σ^2 , the distribution of the number of nerve impulses is asymptotically normal as $t \rightarrow \infty$, with mean $x \sim t/\mu$ and variance

$$s^2 \sim \sigma^2 t/\mu^3 \quad (2.1)$$

If a nerve cell is spontaneously active, the stimulus will start at some random point in an interspike interval and the response is best treated as an equilibrium renewal process (Cox, 1962). For the gamma distribution with parameter r that was considered earlier, it follows from equation (19) of Cox and Miller (1965 p. 343) that the next term in the expansion of s^2 (i.e., the largest error term) is $(1 - r^{-2})/6$. (With an exponential distribution, $r = 1$ and equation (2.1) is an exact expression.) One can treat the data from nerve cells that are not spontaneously active as an ordinary renewal process if the interval distribution for the first impulse is the same as later intervals (Hodgkin, 1948) or a modified renewal process if it is different (Katz, 1950).

The uncertainty of a normal distribution (Shannon, 1948) depends only on its standard deviation s according to the equation

$$H = \log [\sqrt{2\pi}es] \quad (2.2)$$

From equations (2.1) and (2.2) one can calculate the conditional uncertainty produced by a stimulus from the mean μ and standard deviation σ of the interspike interval distributions.

The central limit theorem only proves that the distribution is normal for large numbers of impulses. At the lowest numbers, a Poisson density function, equation (1.14), may provide a better approximation. Then, the conditional uncertainty is

$$\begin{aligned} H(Y/x) &= - \sum_{j=0}^{\infty} p(y_j/x) \log_2 p(y_j/x) \\ &= - \sum_{j=0}^{\infty} [x^j e^{-x}/j!] [j \log_2 x - x \log_2 e - \log_2 (j!)] \\ H(Y/x) &= -x \log_2 (x/e) + \sum_{j=2}^{\infty} [x^j e^{-x}/j!] [\log_2 (j!)] \end{aligned} \quad (2.3)$$

The standard deviation of a Poisson distribution is \sqrt{x} so the error in using equation (2.2) instead of equation (2.3) can be calculated. For $x = 1$ (on average only one impulse is discharged in time t) equation (2.2) gives $\log_2 \sqrt{2\pi}e = 2.04$ bits while equation (2.3) gives 1.88 bits, a difference of only 0.16 bits. As x increases the Poisson distribution approaches a normal distribution and the differences are even less.

(b) *Optimal Stimulus Distributions.* As indicated in section (1 c), the optimal stimulus distribution with small numbers of impulses consists (except at the very ends of the stimulus range) of a set of discrete stimuli whose spacing varies approximately as the standard deviation s of the number of nerve impulses produced in time t . As t increases, the difference between the information one can transmit with such a discrete stimulus distribution and with related continuous distributions decreases. In the limit of large t , the optimal stimulus distribution is in fact a continuous distribution, whose probability density function is given by

$$p(x) = A/s(x) \quad (2.4)$$

where A is a constant such that

$$A^{-1} = \int_{x_{\min}}^{x_{\max}} \frac{dx}{s(x)} \quad (2.5)$$

Equation (2.4) states that the probability density function for the occurrence of particular stimuli must vary inversely as the standard deviation in number of impulses. The standard deviation will be some function $s(x)$ and the factor A merely ensures that $p(x)$ is a proper probability density function; i.e., its integral over the whole stimulus range is one. The proof of equation (2.4) consists of two parts. First, one can show that the inverse relationship between the density of stimuli, $p(x)$, and the variability, $s(x)$, is correct, or, more precisely, that the optimal stimulus distribution consists of a number of equiprobable stimuli whose spacing is proportional to $s(x)$. Then, one can show that the optimal spacing for large t is negligibly small (i.e., the optimal stimulus distribution is a continuous distribution). I will assume throughout that variability is small relative to the range of responses, namely $s(x) \ll x_{\max} - x_{\min}$ for all x . This has the important consequence that effects at the very ends of the stimulus range can be neglected. This assumption becomes increasingly good as t increases since for a nonadapting neuron $x_{\max} - x_{\min} \sim (\nu_{\max} - \nu_{\min})t$ and from equation (2.1), $s(x)$ increases only as \sqrt{t} . Thus, the variability relative to the range of responses decreases as $1/\sqrt{t}$.

The uncertainty of a normal distribution (equation 2.2) is determined by its standard deviation alone. It was shown above that, even when the mean number of nerve impulses is quite small, the conditional uncertainty may approximate closely to the uncertainty of a normal distribution having the same standard deviation. If this is so, one can prove the first part directly by considering a stimulus distribution which consists of a number of equiprobable stimuli whose spacing is proportional to the value of the standard deviation in number of impulses $s(x)$ at each stimulus level. The confusion between stimuli (and hence the information/stimulus) is then a constant, and Fano (1961) showed that this was a condition for maximizing information transmission with a given stimulus set.

To prove the second part, namely that the optimal stimulus spacing becomes

negligibly small as t becomes large, one can transform from a coordinate system x to a new coordinate system x' in which $s(x')$ is approximately constant, by setting $dx' = s(x) dx$. Shannon showed that such a transformation of coordinates will not alter the information transmission. From the first part of the proof, one knows that in the new coordinate system the optimal stimulus distribution consists of a series of equiprobable and, since $s(x')$ is approximately constant, equally spaced stimuli. If there are m stimuli, $p(x_i) = 1/m$ for $i = 1, 2, \dots, m$, the average conditional uncertainty (assuming s large enough that equation (2.2) holds) is

$$H(Y/X) \equiv \sum_{i=1}^m p(x_i) H(Y/x_i) = \log(\sqrt{2\pi e} s)$$

independent of m . Since $H(Y/X)$ is independent of m , and $I \equiv H(Y) - H(Y/X)$, the maximum information will be transmitted where $H(Y)$ is a maximum. This will occur as the number of m equiprobable, equally spaced stimuli becomes very large, because the stimulus distribution and hence the response distribution (variability was assumed small compared to the total stimulus range) will then approach a rectangular distribution. Shannon (1948) proved that a rectangular distribution has the maximum uncertainty of any distribution which is limited to a particular range of values. Thus, the optimal stimulus distribution approaches a continuous distribution although the information transmission is rather insensitive to the exact spacing and, for spacings as large as $3s$ (three standard deviations apart), the number of discriminable categories is only reduced by 1%. To complete the proof, it may be noted that a rectangular distribution in the new coordinate system transforms back to give exactly equation (2.4) in the old coordinate system.

(c) *The Asymptotic Information Capacity for Large t .* Using the results of the last two sections, the information capacity can now be readily derived. Assume once again that t is large enough so that the variability in numbers of impulses is small compared to the total range of responses. Then, the response distribution will be little different from the stimulus distribution, and the response uncertainty is

$$H(Y) \equiv - \int p(y) \log p(y) dy \sim - \int p(x) \log p(x) dx. \quad (2.6)$$

Secondly, the absolute variability in numbers of nerve impulses grows with time, and I will assume that t is large enough so that the conditional uncertainty approaches closely that of a normal distribution (equation 2.2). Then,

$$H(Y/X) \equiv \int p(x) H(Y/x) dx \sim \int p(x) \log [\sqrt{2\pi e} s(x)] dx \quad (2.7)$$

The information transmitted is

$$I \equiv H(Y) - H(Y/X) \sim - \int p(x) \log [\sqrt{2\pi e} p(x) s(x)] dx \quad (2.8)$$

The information capacity is, by definition, the amount of information transmitted when the stimulus distribution is optimal. Equation (2.4) gives the optimal distribution in the limit of large t so

$$C \equiv \max_{p(x)} I \sim - \int_{x_{\min}}^{x_{\max}} p(x) \log [\sqrt{2\pi e} A] dx = - \log [\sqrt{2\pi e} A]$$

Then, from equation (2.5), the basic asymptotic result follows for the information capacity of a neuron using a frequency code

$$C \sim \log \left[\int_{x_{\min}}^{x_{\max}} \frac{dx}{\sqrt{2\pi e} s(x)} \right] \quad (2.9)$$

Substituting the approximate equation (2.1) and the facts that $x = t/\mu$, $dx = -(t/\mu^2) d\mu$, $x_{\max} = t/\mu_0$ and $x_{\min} = t/\mu_1$, a very useful relation is obtained for calculating the information a neuron can transmit in terms of the parameters of the interspike interval distributions.

$$C \sim \log \left[\sqrt{\frac{t}{2\pi e}} \int_{\mu_0}^{\mu_1} \frac{d\mu}{\sqrt{\mu} \sigma} \right] \quad (2.10)$$

σ , the standard deviation of the interval distribution, may vary as a function of the mean interval. The most common relationship found experimentally (Buller, Nicholls and Ström, 1953; Biscoe and Tayler, 1963; Goldberg, Adrian and Smith, 1964; Werner and Mountcastle, 1965; Stein and Matthews, 1965; Biederman-Thorson, 1966; Junge and Moore, 1966; Silk and Stein, 1966) is that the standard deviation increases, at least over parts of its range, as a power function of the mean interval; i.e.,

$$\sigma = b\mu^n \quad (2.11)$$

where b and n are constants and generally $n \geq 1$. In other words, the coefficient of variation σ/μ is a nondecreasing function of μ . Substituting into equation (2.10) and integrating

$$C \sim \log \left[\sqrt{\frac{t}{2\pi e}} \left(\frac{\mu_0^{-k} - \mu_1^{-k}}{bk} \right) \right] \quad (2.12)$$

where $k = n - 0.5$.

Equation (1.16), the asymptotic formula for a gamma distribution is obtained from equation (2.12) by noting that for a gamma distribution $\sigma = \mu/\sqrt{r}$, so $n = 1$, $k = 0.5$, and $b = 1/\sqrt{r}$. It was also assumed in the example that $\mu_0 = 10$ msec and $\mu_1 = 100$ msec.

(d) *Very Regular Neurons.* To obtain equation (2.10) from equation (2.9), the approximate equation (2.1) was used. As indicated earlier, the approximation

will be good if $\sigma^2 t / \mu^3 \gg 1/6$, or

$$t \gg \frac{\mu^3}{6\sigma^2} \quad (2.13)$$

For the gamma distribution example of Fig. 6 *b* where $r = \mu^2/\sigma^2 = 100$, this inequality is valid for a mean interspike interval of 60 msec if

$$t \gg \left(\frac{60}{6}\right) (100) \text{ msec} = 1 \text{ sec.}$$

The approach to the asymptotic formulae will be even slower for more regular neurons. On the other hand, if

$$t \ll \frac{\mu^3}{6\sigma^2} \quad (2.14)$$

the formulae for the completely regular neuron (section 1 *b*) will hold. By retaining an extra constant term in the expansion of s , one can obtain a more accurate asymptotic formula than equation (2.10) for regular neurons

$$C \sim \log \left[\int_{x_{\min}}^{x_{\max}} \frac{dx}{\sqrt{e[1 + 2\pi x/r(x)]}} \right] \quad (2.15)$$

where $r(x) = \mu^2/\sigma^2$ is considered as a function of x and it is assumed that $r(x) > 10$. Integration of this equation for $r(x) = 100$ gives the dotted line on the right of Fig. 6 *b* which considerably improves the fit to the computed (solid) line.

The derivation is as follows: if $r(x)$ is greater than 10, a better approximation to $s(x)$ is given by [see equation (2.1) and the discussion following it]

$$s(x) \sim \sqrt{x/r(x) + 1/6}$$

For the completely regular neuron, $r(x) \rightarrow \infty$ for all x , it would follow from equation (2.2) that the average conditional uncertainty would be $\log \sqrt{2\pi e/6}$. In fact, this slightly overestimates the average conditional uncertainty when a rectangular stimulus distribution is applied. The uncertainty was found to be $1/2$ using natural logarithms, or more generally $\log \sqrt{e}$ [see discussion leading up to equation (1.13)]. Thus, a slightly better approximation for use in informational calculations is given by

$$s(x) \sim \sqrt{\frac{x}{r(x)} + \frac{1}{2\pi}}$$

Substituting this expression into equation (2.9) gives equation (2.15).

For durations above about 50 msec, equation (2.15) may give a more realistic estimate of the maximum information a neuron can transmit than the digital computer values. To attain the computed values, one would need a series of discrete stimuli whose intensities were exactly matched to those of the neuron. As indicated in section (1 *b*), this is highly unlikely under natural conditions, and the optimal stimuli for one neuron will not agree with those for

another with even slightly different firing rates. Equation (2.15) depends only on the "average" way in which variability increases with x , and is therefore appropriate when considering populations of neurons.

(e) *Serial Correlations Between Interspike Intervals.* The derivation of equation (2.10) and equation (2.11) assumed that successive interspike intervals were independent. Serial correlations between adjacent intervals are often found (Kuffler, FitzHugh, and Barlow, 1957; Geisler and Goldberg, 1966) though only the first order serial correlation coefficient ρ_1 may be significantly different from zero (Jansen, Nicolaysen, and Rudjord, 1966). Serial correlation coefficients of all orders are readily accommodated if their values decrease rapidly enough to zero as the order $i \rightarrow \infty$ (Cox, 1962, pp. 104, 134); then equation (2.1) becomes

$$s^2 \sim a\sigma^2 t / \mu^3 \quad (2.16)$$

where $a = 1 + 2 \sum_{i=1}^{\infty} \rho_i$. Substituting equation (2.16) into equation (2.9) produces an expression for the information capacity of a neuron for large t which takes into account serial correlations between successive interspike intervals.

$$C \sim \log \left[\sqrt{\frac{t}{2\pi e}} \int_{\mu_0}^{\mu_1} \frac{d\mu}{\sqrt{\mu^a \sigma}} \right] \quad (2.17)$$

(f) *Nonoptimal Stimulus Distributions.* The proof that equation (2.4) is the optimal distribution makes the basic assumption that the variability in the number of impulses is small compared to the total range of responses. I showed that the assumption became increasingly good as t increases, but to the extent that this assumption is not obeyed, equation (2.4) does not give the optimal stimulus distribution, and the asymptotic formulae, corrected if necessary for extremely regular cells (equation 2.15), tend to underestimate the channel capacity. In addition, the natural or experimental stimulus distributions may not closely approximate the optimal distribution, though one can use the observed stimulus distributions to calculate asymptotic values for information transmission from equation (2.8). With the rectangular stimulus distribution given by equation (1.10), one finds

$$I \sim \log [(x_{\max} - x_{\min}) / \sqrt{2\pi e}] - (x_{\max} - x_{\min})^{-1} \int_{x_{\min}}^{x_{\max}} \log s(x) dx$$

This expression can often be integrated. For example, if one substitutes equations (2.1) and (2.11), one finds after some algebra

$$I \sim \log \left[\sqrt{\frac{t}{2\pi e}} \frac{(\mu_0^{-1} - \mu_1^{-1})}{b} \right] + (n - 1.5) \left(\frac{\mu_0 \log \mu_1}{\mu_1 - \mu_0} - \frac{\mu_1 \log \mu_0}{\mu_1 - \mu_0} - \log e \right) \quad (2.18)$$

Equation (2.18) reduces to equation (2.12) if $n = 1.5$ (a rectangular distribution is then the optimal distribution since $s = t\sigma^2/\mu^3 = bt$, independent of x). When $\mu_1 \rightarrow \infty$ ($\nu_{\min} \rightarrow 0$),

$$I \sim \log \left[\sqrt{\frac{t}{2\pi}} \frac{e^{1-n}}{b\mu_0^{n-1}} \right]$$

For a gamma distribution ($n = 1$) and assuming $\mu_0 = 10$ msec and $\mu_1 = 100$ msec, equation (2.18) predicts a value for I independent of the value of r which is 0.08 bits less than the information capacity given by (2.12). Thus, application of a rectangular (rather than an optimal) stimulus distribution would only reduce the number of discriminable categories by 6%, often a negligible amount.

(g) *Stationary and Nonstationary Discharge Patterns.* Each of the asymptotic formulae I have derived (equations 2.10, 2.12, 2.15, 2.17, and 2.18) has the property that for large t ,

$$I \sim \log \sqrt{t/T} \quad (2.19)$$

where T is a constant. This is a direct consequence of the central limit theorem and will hold whenever the discharge is stationary and the interspike interval has a finite variance ($\sigma^2 < \infty$).

The central limit theorem insures that $s(x)$ increases as \sqrt{t} , so one can write $s(x) = s(\nu)\sqrt{t}$, where $x = \nu t$. Similarly, if the discharge is stationary, $p(x) = p(\nu)/t$ and $dx = t d\nu$, so equation (2.8) becomes

$$\begin{aligned} I &\sim - \int p(\nu) \log [\sqrt{2\pi e} p(\nu)s(\nu)/\sqrt{t}] d\nu \\ &= \log \sqrt{\frac{t}{2\pi e}} - \int p(\nu) \log [p(\nu)s(\nu)] d\nu \end{aligned}$$

The integral on the right is independent of time and so the equation is of the form of equation (2.19) where

$$T = 2\pi e^{1+B} \quad (2.20)$$

and $B = \{ \int p(\nu) \ln [p(\nu)s(\nu)] d\nu \}^2$

All the limiting relations, therefore, are parallel with a slope of $1/2$ on a plot of C versus $\log t$. Measurement of four times as long a discharge will yield one bit more information or twice the number of discriminable categories. The variability or regularity of discharge, and the extent to which the optimal stimulus distribution is approached, only shift the line along the time axis.

If the impulse frequency adapts to a maintained stimulus (the discharge is then

not stationary), the advantage of longer stimuli will be less significant since the number of impulses will increase less quickly with time. If there is any random element linking the natural stimulus to the discharge which causes the mean frequency to fluctuate from stimulus to stimulus with a variance $f^2(\nu)$, then the total variance in number of impulses will have two components

$$s^2(x) = ts^2(\nu) + [tf(\nu)]^2$$

and

$$C \sim \int_{\nu_{\min}}^{\nu_{\max}} \frac{d\nu}{\sqrt{2\pi e f(\nu)} \left[\sqrt{1 + s^2(\nu)/(t f^2(\nu))} \right]} \quad (2.21)$$

For t sufficiently large, the square root on the right must approach unity and the channel capacity is then independent of time. In other words, the maximum amount of information that can be transmitted about a single stimulus via a frequency code is ultimately limited by the reproducibility of the mean frequency of nerve impulses generated by the stimulus.

(h) *More Than One Cell.* Stimuli do not normally affect single nerve cells, but populations of cells. Stronger stimuli will not only increase the frequency of cells that are already active, but may also recruit new cells. If the important variable is the total number of nerve impulses discharged by N cells in time t , one can write from standard formulae (Feller, 1957)

$$x = \sum_{i=1}^N x_i$$

$$s^2 = \sum_{i=1}^N s_i^2 + 2 \sum_{i,j} \rho_{ij}$$

where x_i and s_i^2 are the mean and variance in the number of nerve impulses discharged by the i th cell in time t . ρ_{ij} is the correlation coefficient between the number discharged by the i th and the j th cell and the second sum extends over all pairs of i and j for $j = 2$ up to N where $i < j$. In the special example of N identical, independent nerve cells, it is easily shown that an equation such as equation (2.10) would become

$$C \sim \log \left[\sqrt{\frac{Nt}{2\pi e}} \int_{\mu_0}^{\mu_s} \frac{d\mu}{\sqrt{\mu\sigma}} \right] \quad (2.22)$$

The asymptotic information capacity for a given value of t is increased by $\frac{1}{2} \log N$. Alternatively, one could transmit the same information in a fraction t/N of the

time. (In fact, one would generally be able to transmit rather more as the asymptotic formulae tend to underestimate information transmission for short t .) If the neurons are independent, but not identical, the central limit theorem again insures that the number N affects the information transmission as in equation (2.22). In practical situations, the extent to which information transmission can be enhanced by summation of numbers of cells will be limited by any underlying correlation between neurons; i.e., the extent to which stimulus independent variations in the frequency of one neuron are correlated with those of the rest of the population.

These results apply only to functionally similar groups of neurons. An obvious example of a functionally similar group is the alpha motoneurons which innervate a synergic group of muscles, and an equally obvious counter-example is the entire population of alpha motoneurons in a ventral root. A stimulus may excite flexor motoneurons and inhibit extensor motoneurons without changing the total number of nerve impulses very much. However, if the two groups of motoneurons are analyzed separately, a considerable amount of information may be available about the strength of the flexor reflex stimulus.

Another clear example of a functional group is those neurons which excite a single higher order neuron. The information about a stimulus contained in the discharge of higher order neurons can be analyzed and compared with that of lower order cells. In this way one can study the flow of information in the nervous system, but one must be careful to use appropriate stimulus patterns for each cell. The additional processing at each synapse may soon lead to rather complex pattern analysis (Powell and Mountcastle, 1959; Hubel and Wiesel, 1963; Oonishi and Katsuki, 1965), and the total information present at each level is the sum of the number of bits of information available about each pattern that is analyzed independently.

DISCUSSION

In Fig. 7 the results for neuronal information transmission using a frequency code are summarized and compared schematically with those using other codes. (In this figure the small periodicities which are evident in Figs. 2 and 6 were neglected, and it was assumed that the neuron was discharging spontaneously.) All the codes give similar values for very short times, when the probability of an impulse is much less than one. The neuron can then be considered as a binary channel and its capacity is given by equation (1.1). However, as the stimulus duration increases and multiple responses occur more frequently, the scope for efficient coding expands enormously. The properties of a nerve cell as a communications channel can be summarized in a plot of information capacity against the duration of a maintained stimulus. Curve *a* in Fig. 7 represents the capacity of a randomly discharging cell which uses a frequency code. If the capacity is measured in bits of information and the time scale is logarithmic with each unit representing a doubling of stimulus duration, then equation (1.15) indicates that the slope of such a plot will approach $\frac{1}{2}$ at long times. This

is a general limiting result for frequency coding (section 2 g) if (1) the discharge is stationary and (2) the stimulus generates an exactly reproducible mean impulse frequency. (There are other conditions which are unlikely to cause trouble in practice; e.g., the variance of the interspike interval distribution must be finite.) From the equation (2.12) or (2.17) one can calculate the asymptotic information capacity directly from the parameters of the interspike interval distributions which are now commonly measured experimentally.

If the discharge is sufficiently regular (curve *b*), the approach to the limit will be slow (section 2 d), and the slope on an information versus log time plot may have a slope of one until times long enough that the inequality (equation 2.14) no longer

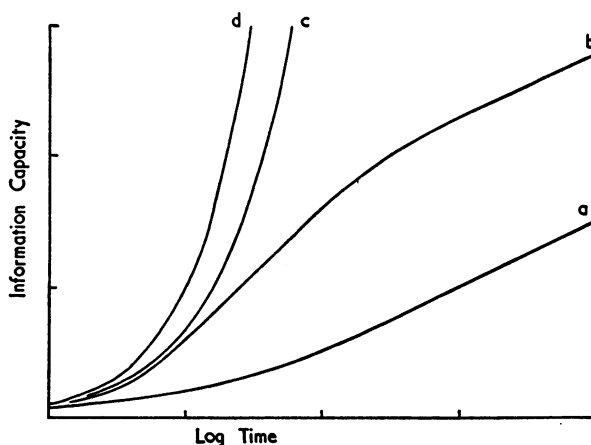


FIGURE 7 Schematic representation of the information capacity as a function of stimulus duration for a neuron, (a) discharging randomly and using a frequency code, (b) discharging fairly regularly and using a frequency code, (c) using a binary pulse code, and (d) using an interval code. Explanation in text.

holds. To the extent that the conditions above are not satisfied, the capacity will be further limited at long times and may approach a constant value (section 2 g). This effect was noted experimentally by Werner and Mountcastle (1965).

Curve *c* in Fig. 7 represents the capacity using a binary pulse code. With a binary code, a constant number of bits/second can be transmitted so the capacity increases linearly with time, and exponentially when plotted against $\log t$. Curve *d* in Fig. 7 represents the information transmitted using an interval code. Mackay and McCulloch (1952) showed that using the length of each interval to encode information might further increase the number of bits/second transmitted and hence curve *d* is a larger exponential in this figure.

Mackay and McCulloch (1952, p. 134) were careful to point out that their calculations "must not be taken to imply a belief that either binary coding or pulse interval coding in the communications engineer's sense is the mode of operation of

the central nervous system." In fact, they noted that "there has been steady accumulation of other evidence which has been adduced in favor of models employing frequency modulation," and that "the statistical nature of nervous activity must preclude anything approaching realization in practice of the potential information capacity of either mechanism." Yet, in the 15 yr since their article, there has been to my knowledge no comparable analysis of the information that can be transmitted by neurons which use a frequency code and are subject to variability. I have therefore concentrated in this paper on frequency coding and the effects of variability, though I would not claim that all neurons use frequency coding. The nervous system can locate a sound source accurately by using mutually inhibitory connections between neurons on the two sides of the body to detect the difference in time of arrival of the sound at the two ears (Hall, 1964). Thus, some precise temporal information is available in addition to that contained in the mean frequency of firing. Hall notes that "these cells may be thought of as logical transducers that convert stimulus differences at the two ears into patterns of response activity which can be 'read' by higher neural centers." In other words, the output from this group of cells is in terms of total numbers of nerve impulses. Thus, the results of this paper are applicable, and the amount of precise temporal information retained should be calculable (though, at least with brief "click" stimuli, recruitment of more nerve cells rather than temporal summation is used exclusively). As indicated in section (2 *h*), there is a basic trading relationship between stimulus duration and the number of functionally similar cells. If the stimulus is shortened by a given factor, the same amount of information can often be transmitted by multiplying the number of nerve cells by the same factor

Except in section (2 *h*), relatively little has been said about the way in which information is decoded or, in other words, the extent to which it is available to higher order neurons. The formulae give values only for the maximum information that can be obtained from knowing the total number of impulses that were discharged in a given period of time. Many of these impulses will be ineffective in synaptic transmission for several reasons. First, the higher order neurons will be refractory a certain part of the time. Secondly, if a neuron is discharging slowly, the effects of many impulses will decay away during an interspike interval and not contribute directly to action potential initiations. Finally, inhibitory inputs which cause membrane hyperpolarization (or subtract from membrane depolarization) will increase variability at a given mean frequency (Stein, 1965) and hence decrease information transmission. However, these examples illustrate the general rule (Ash, 1965 p. 85) that data processing cannot increase the amount of information. Data processing merely transforms it into a more useful form, possibly at the cost of considerable loss of information. This loss in neuronal data processing can be assessed experimentally since the same formulae apply to all nerve cells which use frequency coding. Thus, one can determine how much information higher order cells retain about the stimulus.

Finally, a comment is needed on the magnitude of the differences between the information capacities of systems using frequency coding, and those using binary or interval coding. Calculations from preliminary experimental data (Stein and Matthews, 1965) indicate that even extremely regular muscle spindle afferent fibers transmit, at most, about a half dozen bits of information about a maintained stimulus lasting a second (see also Werner and Mountcastle, 1965), not the hundreds or thousands which binary or interval codes predict. However, it must be recalled that the capacities calculated here represent the amount of information transmitted about a single stimulus as a function of its duration. To convert these figures to bits/second for direct comparison with earlier estimates, it is necessary to specify the stimulus repetition rate. The maximum repetition rate that a nerve cell can follow adequately must be determined experimentally. One cannot assume that, because one bit of information is transmitted about a stimulus lasting t milliseconds, $1000/t$ bits/sec can be transmitted about stimuli presented every t milliseconds, unless it has been shown experimentally that the responses to successive stimuli are independent (cf. Jacobsen, 1951). The maximum allowable repetition rate will decrease at higher levels of the nervous system since synaptic transmission tends to integrate (excitatory) presynaptic inputs over a period of time (Maffei and Rizzolatti, 1967).

Similarly, to specify the capacity of a system or whole organism, one cannot multiply the bits/second transmitted by one cell by the number of cells unless one has shown that adjacent cells or groups of cells function as separate channels. At higher levels of the nervous system, spatial summation of the impulses generated by large numbers of nerve cells may severely limit the number of such separate channels, and at the level of decision making, there is surprisingly little independence, even between the two sides of the body (Broadbent, 1958) or different sensory systems (Davis, 1957). According to equation (2.22) the same stimulus applied simultaneously to 1000 independent, functionally similar cells would provide only 5 bits more information instead of 1000 times as much, which would result if each cell were used as a separate channel. These differences should help tremendously in bringing estimates of the information capacity of single cells and systems of cells into line with those of human information capacities which were mentioned in the Introduction. The reasons for the apparently vastly inefficient means of transmitting information have also been considered and might be summarized by saying that the aim of the biological organism is maximum survival, not information transmission. However, once one understands how the nervous system codes the patterns which are necessary for its survival, information theory should provide a very useful way to measure efficiency and to compare different sensory modalities and different levels of the nervous system.

Dr. F. C. Marriott, Dr. D. Noble, Mr. K. G. Pearson and Mr. R. W. Tsien made helpful suggestions.

This work was supported in part by a U. S. Public Health Service Fellowship 1-F2-GM-29-201-01.

Received for publication 27 March 1967.

REFERENCES

- ADRIAN, E. D. 1932. *The Mechanism of Nervous Action*. University of Pennsylvania Press, Philadelphia.
- ARMETT, C. J., J. A. B. GRAY, R. W. HUNSPERGER, and S. LAL. 1962. *J. Physiol.*, (London). **164**:395.
- ASH, R. 1965. *Information Theory*. John Wiley & Sons, Inc., New York.
- BARLOW, H. B. 1963. *Kybernetik*. **2**:1.
- BIEDERMAN-THORSON, M. 1966. *J. Gen. Physiol.* **49**:597.
- BISCOE, T. J., and A. TAYLOR. 1963. *J. Physiol.*, (London). **168**:332.
- BROADBENT, D. E. 1958. *Perception and Communication*. Pergamon Press, Ltd., Oxford.
- BULLER, A. J., J. G. NICHOLLS, and G. STRÖM. 1953. *J. Physiol.*, (London). **122**:409.
- CALVIN, W. H., and STEVENS, C. F. 1966. Abstracts of the 2nd International Biophysics Congress, Vienna, Austria. 424.
- COX, D. R. 1962. *Renewal Theory*. Methuen & Co. Ltd., London.
- COX, D. R., and H. D. MILLER. 1965. *The Theory of Stochastic Processes*. Methuen & Co. Ltd., London.
- DAVIS, R. 1957. *Quart. J. Exptl. Psychol.* **11**:211.
- FANO, R. M. 1961. *Transmission of Information*. The M. I. T. Press, Cambridge.
- FELLER, W. 1957. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, Inc., New York. 1.
- FITZHUGH, R. 1957. *J. Gen. Physiol.* **40**:925.
- GEISLER, C. D., and J. M. GOLDBERG. 1966. *Biophys. J.* **6**:53.
- GOLDBERG, J. M., H. O. ADRIAN, and F. D. SMITH. 1964. *J. Neurophysiol.* **27**:706.
- GRÜSSER, O. J. 1962. *Kybernetik*. **1**:209.
- HALL, J. L. 1964. *Mass. Inst. Technol. Res. Lab. Electron. Tech. Rept.* **416**.
- HODGKIN, A. L. 1948. *J. Physiol.*, (London). **107**:165.
- HUBEL, D. H., and T. N. WIESEL. 1963. *J. Physiol.*, (London). **165**:559.
- JACOBSEN, H. 1950. *Science*. **112**:143.
- JACOBSEN, H. 1951. *Science*. **113**:292.
- JANSEN, J. K. S., K. NICOLAYSEN, and T. RUDJORD. 1966. *J. Neurophysiol.* **29**:1061.
- JUNGE, D., and G. P. MOORE. 1966. *Biophys. J.* **6**:411.
- KATZ, B. 1950. *J. Physiol.*, (London). **111**:248.
- KUFFLER, S. W., R. FITZHUGH, and H. B. BARLOW. 1957. *J. Gen. Physiol.* **40**:683.
- MACKAY, D. M., and W. S. MCCULLOCH. 1952. *Bull. Math. Biophys.* **14**:127.
- MAFFEI, L. and G. RIZZOLATTI. 1967. *J. Neurophysiol.* **30**:333.
- ONISHI, S., and Y. KATSUKI. 1965. *Japan. J. Physiol.* **15**:342.
- POWELL, T., and V. B. MOUNTCASTLE. 1959. *Bull. Johns Hopkins Hosp.* **3**:108.
- QUASTLER, H. 1956. In *3rd Symposium on Information Theory*. C. Cherry, editor. Butterworth & Co. Ltd., London. 361.
- RAPOPORT, A., and W. J. HORVATH. 1960. *Inform. Control*. **3**:335.
- SHANNON, C. E. 1948. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana. Reprinted 1962.
- SILK, N., and R. B. STEIN. 1966. *J. Physiol.*, (London). **186**:40P.
- STEIN, R. B. 1965. *Biophys. J.* **5**:173.
- STEIN, R. B. 1967. *Biophys. J.* **7**:37.
- STEIN, R. B., and P. B. C. MATTHEWS. 1965. *Nature*. **208**:1217.
- WERNER, G., and V. B. MOUNTCASTLE. 1965. *J. Neurophysiol.* **28**:359.